5th International Conference on Corpus Linguistics (CILC2013)

Dialectal Variation in a Nineteenth-Century Catalan Grammar Corpus

Maria-Pilar Perea[a]*, Manel Sifre[b]

[a]Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007 Barcelona, Spain
[b]Universitat Jaume I, Av. Vicent Sos Baynat s/n, 12071 Castelló de la Plana, Spain

Abstract

One of the main objectives currently developed by the LEXDIALGRAM project is the creation of a grammar corpus obtained from digitizing Catalan dialect works that were written during the nineteenth century. Digitized materials are included in an internet web-portal (http://www.ub.edu/lexdialgram/) in order to promote their access and dissemination. Currently eight grammars and two orthographies corresponding to the Valencian, Majorcan, Menorcan, Algherese and Rossellonés varieties are included. The texts were processed using a computer tool that has facilitated access to materials through simple and complex queries. Computerization has helped to develop studies based on the analysis of dialectal language variation, from both a synchronic and diachronic viewpoint. The corpus allows us to develop sociolinguistic analysis and issues arising from language contact, since in many cases the works are bilingual or trilingual. This paper is divided into two parts: the first describes the corpus and the computer tool used to process the Catalan grammars; the second examines two grammatical phenomena. Altogether, it is a sample of the potential of this data as a resource to describe linguistic variety in an area and in a given time period, and the possibility of accomplishing comparative dialectal studies.

Keywords: Corpora; grammar; Catalan; dialects

* Corresponding author. Tel.: +0-0034-934035632; fax: +0-0034-934035698.
  E-mail address: mpilar.perea@ub.edu.

1. Introduction

The Web-Portal for the study of nineteenth-century Catalan dialects (http://www.ub.edu/lexdialgram/), created as part of the LEXDIALGRAM project, brings together a variety of scholarly works on the subject and classifies them under the headings of lexicography and grammar: dictionaries, lexicons and vocabularies on the one hand, and grammars and spelling guides on the other. Its main objectives are to bring these works to a broader public and to promote studies that analyse the state of the Catalan language and its dialects in the nineteenth century. A range of procedures of varying degrees of complexity have been applied to aid consultation of these materials, including digitization, computerization and design applications; in the future, we hope to add mapping as well.

A database of the lexicons has already been created which contains all the words that appear in the selected works (cf. Perea 2012a & 2012b). As for the grammars, the aim is to build a corpus resulting from the digitization of ten works written during the nineteenth century: eight grammars and two spelling guides, covering the dialectal varieties of Valencia, Majorca, Minorca, Roussillon and Alghero. The texts from the grammars have been processed using a computer application that allows users to access the materials through a variety of search queries. In fact, computerization has made it possible to carry out studies of both synchronic and diachronic dialectal variation. The grammar corpus is also of use in studies in the areas of sociolinguistics and linguistic standardization, and because many of the works are bilingual or trilingual, the corpus is especially relevant to issues of language contact.

This presentation is divided into two parts. The first describes the corpus and the computerization of the materials. The second examines two grammatical phenomena and stresses the potential of the data as a resource for describing and comparing the linguistic variety in a given area and time period.

2. Corpus description

2.1. Chronological context

The nineteenth century has been chosen as the chronological context for this study of Catalan dialects because a great many of the dictionaries and grammars published during that century centred on a specific geographical location – Majorca and Valencia are two examples among many. However, dialectal studies did not really emerge as an area of research until the end of the century. Consequently, in order to understand the situation of dialects in the nineteenth century, the materials compiled represent an important source of information.

There are three main reasons for the preponderance of studies of this kind in the period under analysis. The first is the regional fragmentation of the Catalan-speaking territories and the gradual loss of the sense of linguistic unity. This explains the proliferation of dictionaries and grammars in Catalan, Valencian, Majorcan, Minorcan and so on, each produced by local writers focusing on their own dialect and in isolation from the others.

The second factor is the role played in the many vocabularies and grammars of this period by Castilian and, to a lesser extent, by Latin, as the languages used to translate Catalan. Indeed, most of the Catalan grammars were written in Castilian (with the exceptions of the ones published in Alghero [in Italian] and in Perpignan [in French]). Our corpus contains only one work written in Catalan: the *Principis de la lectura menorquina* by Juan Ramis i Ramis.

The third factor leading to the emergence of studies of dialects is the ideological influence of Romanticism and the Renaixença (the Catalan "Renaissance"), which encouraged intellectuals of the nineteenth century to recover the roots of their country through the compilation of folkloric data and language. Many journals of the time, especially those edited by hiking clubs, published studies of traditional life and provide interesting lists of words and grammatical features specific to a particular town or area.

2.2. The grammar corpus

The grammar corpus of LEXDIALGRAM consists of ten works: eight grammars and two spelling guides. Since most of the grammars include a section on spelling, it was considered appropriate to include these books in the corpus. The geographical areas represented are Valencia, Majorca, Minorca, Roussillon and Alghero.

- Grammars

1. Amengual, Joan Josep (1836) *Gramàtica de la llengua mallorquina*, 2a ed., Palma: Imp. de P. J. Gelabert, 1872, 226 p.

2. Pere Puiggarí (1852) *Grammaire catalane-française, à l'usage des français...*, Perpignan: J. B, Alzine, 134 p.

3. Julio Soler (1858) *Gramática de la lengua menorquina*, Maó: Imprenta de D. J. Fábregues y Pascual, 128 p.

4. Josep Nebot i Pérez (1894) *Apuntes para una gramática valenciana popular*, València: Impremta Ripollés.

5. Joan Pais i Melis (1899) *Grammatica del dialetto moderno d'Alghero*, 1899 (edición de Pasqual Scanu, 1970).

6. Foulché-Delbosc, Raymond (1902) *Abrégé de grammaire catalane*, Barcelone: Imprimerie et Librairie "L'Avenç".

7. Palomba, Giovanni (1906) *Grammatica del dialetto algherese odierno*, Sasari: Tipografia G. Montorsi.

8. Tomás Forteza (1915) *Gramática de la lengua catalana*, finished in 1898, published in 1915, Palma: Esc. tipogr. prov., 563 p.

- Spelling guides

1. Un MAHONÉS [Joan Ramis i Ramis] (1804) *Principis de la lectura menorquina*, Mahò: Impr. de la Vda. de Fàbregues.

2. F. A. M. S. M. (1812) *Nueva ortografia de la lengua mallorquina*, Palma: Impr. de Sebastian Garcia.

2.3. Methodology

These materials have been scanned, transcribed and stored in a user-friendly computer application. The index and the subdivisions of each work can be consulted. As shown in Figure 1, the basic information on the work is entered on the screen; then each chapter and subchapter are examined, and finally the content of each paragraph is reproduced.

Thus, the materials can be easily accessed and their structures and contents can be compared. The location of the particular feature in the grammar is recorded (chapter, section, page, paragraph), as is the phenomenon to which it refers, especially if it is to do with phonetics. Tomas Forteza's *Gramática de la lengua catalana*, for example, contains a great deal of phonetic data. The Majorcan scholar refers to the phenomenon of elision of consonants (Table 1) in some areas of Catalonia, which in turn is compared with its presence in the Balearic area. Its appearance confirms that the weakness of the schwa is not the reason for its loss:

Table 1. Sample of phonetic phenomenon: Elision of consonants in Majorcan.

| Location	Text	Phenomenon
Capítulo X. [Figuras de dicción. Síncopa]. § 396 <Chapter X. Figures of speech. Syncopation. § 396>	"Suprímese la *d* cuando la precede una consonante y la sigue la *r; v.* gr.; *penre* (pendre), *venre* (vendre), *entenre* (entendre), *porá* (podrá); los valencianos la suprimen entre dos vocales que pueden formar diptongo, v. gr. *caíres* (cadires), *guisaet* (guisadet), *vegaes* (vegades), *bofetaes* (bofetades), *garrotaes* (garrotades), etc. especialmente en los acabados en *ador, adora*, como: *cassaor* (cassador), *jarraor* (xarrador), *pescaora* (pescadora), *treballaora* (treballadora), etc." ('The letter *d* is deleted when it is preceded by a consonant and is followed by *r*, e.g.; v. gr.; *penre* (pendre) 'to take', *venre* (vendre) 'to sell', *entenre* (entendre) 'to understand', *porá* (podrá) 'he will be able'; Valencian speakers delete it between two vowels that can create a diphthong, e.g. *caíres* (cadires) 'chairs', *guisaet* (guisadet) 'stew', *vegaes* (vegades) 'times', *bofetaes* (bofetades) 'slaps', *garrotaes* (garrotades) 'beats with a stick', etc. especially words ending in *ador, adora*, such as: *cassaor* (cassador) 'hunter', *jarraor* (xarrador) 'talkative man', *pescaora* (pescadora) 'fisherwoman', *treballaora* (treballadora) 'working woman', etc.').	Elision of consonants
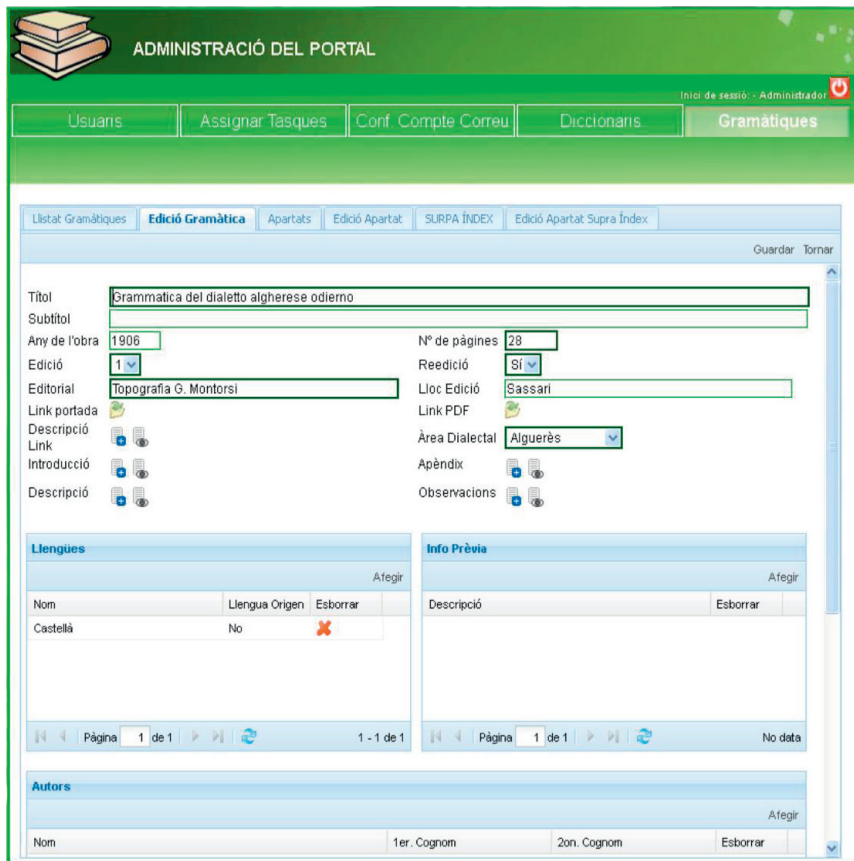
Fig. 1. The screen for entering grammatical data

Some of these phenomena can also be found in other grammars, for example, in José Nebot y Pérez's *Apuntes para una Gramática Valenciana* (Table 2):

Table 2. Sample of a phonetic phenomenon: Elision of consonants in Valencian.

Location	Text	Phenomenon
Capítulo XI [Figuras de dicción], p. 115 'Figures of speech'	Pero donde más se emplean estas figuras es en las palabras terminadas en ador, *adora* y *esa*, en las cuales casi puede asegurarse en absoluto que se suprime la d ó la s; así decimos, *mocaor* y *llauraora* en vez de *mocador* y *llauradora*, *perea* y *riquea* por *peresa* y *riquesa*. ('But where these figures are used most is in words ending in *ador*, *adora* and *esa*, in which *s* or *d* is deleted; so we say, *mocaor* 'handkerchief' and *llauraora* 'farmer woman' instead of *mocador* and *llauraora*, *perea* 'laziness' and *riquea* 'wealth' instead of *peresa* and *riquesa*').	Elision of consonants

Note that although these descriptions represent the same phenomenon, they are included in different sections of the two grammars. For this reason, the program includes the possibility to make a classification by linking descriptions of identical or similar phenomena.

2.4. Difficulties in the computerization of the content of the grammars

The introduction of the contents of the grammars in the computer program has encountered a variety of difficulties in the following areas:

a) Language: The title of the work shows the language used in the grammar: usually Castilian, and also French and Italian. This accounts for the special emphasis given to the examples, as they are the only information that is written in Catalan.

b) Spelling: Problems are created by non-standard spellings.

c) Non-grammar content: Besides information on grammar, some works also contain lexical aspects such as:

1) Lists of words with a translation into Castilian. In Table 3, words belonging to a particular semantic field, "food", in the *Gramática de la lengua menorquina* by Juli Soler ([Apéndice. Voces usuales 'Appendix. Usual words'] p. 104). In several cases, both the Majorcan words and the corresponding Spanish words raise phonetic issues.

Table 3. A sample of list of words in the Soler's *Gramática*.

Catalan	Castilian
cuera	cuchara 'spoon'
sacatrapos	tirabuzon 'corkscrew'
ráveg	rábano 'radish'
xúa	tocino 'bacon'
butifarra	morcilla 'black sausage'
cobrombos	pepinos 'cucumber'
berenar	almorzar 'breakfast'
coc	panecillo francés 'roll'
ciurons	garbanzos 'chickpeas'
pebre bo	pimienta 'pepper'
pebre vermell	pimiento colorado 'red pepper'
perdiu	perdiz 'partridge'

2) Lists of words with the corresponding definition, as shown in the *Nueva ortografía de la lengua mallorquina, explicada en español para su más fácil inteligencia*, by F. A. M. S. M. (ARRASTRÁD. Qui fá mála vida. *DÓNA ARRASTRÁDA*. Lo matéx que prostituta, raméra; ÁSCO. Alteraciò d'es ventrey causáda per se repugnáncia que's té à álguna cósa, qui provóca à vómit. Es vulgo diu: OY. ('DRAGGED: Person who lead a bad life. *WOMAN DRAGGED ALONG*. The same as a prostitute, whore; REVULSION. Alteration of the belly caused by the repugnance one feels towards something that causes vomiting. People say: OY').

In this case, the title of the section in which the words are defined indicates the abundant presence of Spanish loanwords, and also the author's opinion regarding their use: "They are very useful for youth who wish to speak with beauty and energy". These works also express a specific language ideology, characterized variously by prejudice, diglossia, linguistic insecurity, pride or self-hatred.

These data are included in the lexicon database rather than the grammar corpus.

2.5. The value of the examples

In most grammars, the only data given in Catalan are the examples. To maintain this information, in the text examples are inserted between symbols << >>, so that they can be subsequently extracted, categorized and linked to a set of phonetically, morphologically and syntactically predetermined descriptors. The descriptors corresponding to phonetic phenomena are listed below:

Descriptors [Phonetics]:

a) Stress (and stress shift)

b) Consonants: 1. affrication, 2. 'apitxament' (devoicing of voiced alveolars and palatals), 3. approximants, 4. assimilation, 5. loss of -*r*, 6. deaffrication, 7. depalatalización, 8. dissimilation, 9. elision, 10. stops, 11. epenthesis, 12. labiodental fricative ([v]), 13. fricatives, 14. gemination; 15. consonant clusters in final position; 16. consonant clusters, 17. *yeísmo*, 18. sound change, 19. palatalization, 20. palatalization of *l*, 21. palatalization of velars, 22. prothesis, 23. rhotacized vowels, 24. glides, 25. elision, 26. voicing, 27. vocalization of *l*.

c) Sound contact: 1. elision.

d) Vowels: 1. diphthongization, 2. diphthongs, 3. dissimilation, 4. elision, 5. epenthesis, 6. vowel harmony, 7. sound change, 8. monophthongization, 9. palatalization, 10. vowel reduction, 11. prosthesis, 12. semiconsonants, 13. closed vowels.

e) Unstressed Vowels: 1. loss of pretonic schwa, 2. proparoxytones final vowels, 3. final vowels.

f) Stressed vowels.

When a linguistic phenomenon appears in the grammars but is not included in the default list, it can be added and selected if it reappears.

2.6. Quantitative analysis

The contents of grammars grouped according to their language (Castilian, Catalan, Italian or French) are also processed through the program Wordsmith 6.0 for quantitative analysis of the words in the corpus.

The program has created a list of 33,002 different words sorted according to frequency (ranging from the most frequent, the preposition DE 'of' = freq. 10,564, to the least frequent, the verb ZURZIR 'darning' = freq. 1). The selection of the 300 most frequent words indicates which terms or areas are preeminent in the grammars. The ten most frequent words are the prepositions *de*, *en* and *a*; the articles *la*, *el* and *los*; the conjunctions *y* and *que*; the pronoun *se* and the adverb *no*. Ninety-eight other words are related to grammar issues, or sixty-seven if we group together the words that belong to a same lemma (i.e., adjetivo/s 'adjective/s' (250); adverbio/s 'adverb/s' (201), etc.). In case of ambiguity, concordances make it easy to determine the aspect of grammar to which the word refers.

The resulting order, based on the absolute frequency of the words, shows a number of grammar issues (Table 4) in the works studied. The list of most frequent words includes concepts related to language, Catalan, but also to other languages. A first glance shows that the study of the verb and the terminology associated with it is one of the major issues, although there are also references to other categories (nouns, adjectives and verbs) and other parts of speech (pronunciation, sound, compounds). The words *example/examples* are frequent, and two reference works are included in Forteza's *Gramática* ("Aguiló" referring to the works written by the Majorcan scholar Marià Aguiló, and "conq", which refers to the *Crònica* by Jaume I).

Table 4. Most frequent words in the grammar corpus.

LENGUA/S	791	DESINENCIA/S	218	CONSONANTE	123
CAT/CATALÁN/ANA/ CATALUÑA	651	PRIMERA	218	TERCERA	117
LATÍN/O/INA/S	536	SONIDO	218	SIGNIFICACIÓN	113
VERBO/S	510	SING/SINGULAR	216	FRASES	110
CASTELLANO/A	424	TERMINACIÓN/TERMINADOS	216	ACENTO	109
NOMBRE/S	363	INDICATIVO	207	FRANCÉS	108
PRESENTE	355	ADVERBIO/S	201	FORMACIÓN	106
VOZ/VOCES	349	VALENCIA/VALENCIANO	197	GERUNDIO	103
USA/USO	346	LETRA/S	187	AGUILÓ	100
VOCAL/ES	338	IMPERFECTO	185	ESCRITORES	99
MALL/ MALLORCA	329	PARTICIPIO	181	FUTURO	98
PALABRA/S	308	PRONOMBRE/S	171	TÓNICA	95
EJEMPLO/S	301	SUBJUNTIVO	170	PERFECTO	93
PRETÉRITO	275	CONJUGACIÓN	165	POPULAR	91
FORMAS	254	SEGUNDA	164	CONQ[ueridor]	89
PERSONA/S	254	LENGUAJE	161	FEMENINO	89
PLUR/ PLURAL	253	COMPUESTOS	156	SÍLABA	88
TIEMPO/S	252	INFINITIVO	139	ANTIGUOS	85
ADJETIVO/S	250	REGLA	139	DICCIÓN	85
SUSTANTIVO/S	248	VULGAR	137	DERIVADOS	81
CASO/S	246	ARTÍCULO	135	MODO	81
PRONUNCIA/ PRONUNCIACIÓN	246	IMPERATIVO	124	PREPOSICIÓN	80

The quantitative study of the words in the grammars also provides information on the dominant language areas or concepts that the author wishes to highlight due to their complexity or interest.

3. Grammatical analysis

There are literally hundreds of aspects that could be taken into consideration, because they all contribute both to understanding the facts of the language and to finding the best methodology for approaching linguistic science. For this article a representative selection has been made. Regarding the facts of the language, we have chosen some morphological aspects, which will be developed later; and regarding the methodology, a fragment from Forteza's *Gramática* (§ 218, p. 158), because of the admirable lucidity of his approach: "Los gramáticos modernos han formado del sustantivo y el adjetivo dos partes de la oración distintas y separadas. Esta separación ni parece bastante fundada, ni conduce á simplificar el estudio; condiciones que deberían exigirse á toda innovación por sencilla que fuese, á trueque de las dificultades que engendra." ('Modern grammarians have made the noun and the adjective two distinct parts of the sentence. This division does not seem to be well founded, and does not help to simplify the study – conditions that should be required of any simple innovation, in spite of the difficulties that may arise').

The consensus of current linguistic science has moved away from the traditional grammar view according to which the noun is a unique category with two divisions (a noun and an adjective). This is true. But Forteza's proposal requires that innovations in scientific knowledge should fulfil are self-evident, once we have understood them. A well-founded proposal involves knowing its precedents, noting any deficiencies, and proposing an alternative. And the alternative has to overcome the deficiency and, if the object of study allows it, has to be simpler. With this, it is clear that Forteza was, aware, either directly or through his practice as grammarian, of the traditional methodological principle of Occam's razor.

As for the morphological aspects, we have focused on nominal inflection and on the appreciative derivation (mainly augmentative and diminutive), which is relatively close to it. The corpus of grammars has few definitions of gender and number; in fact, the only one to provide definitions of this kind is Amengual, which is also the oldest. His idea of number is sound (p. 4): "Número en la palabra, es la espresión de la idea accesoria de la cantidad, añadida á la idea principal de la palabra misma." ('Number in a word is the expression of the accessory idea of amount, added tothe main idea of the word itself').

Number is, therefore, an idea of the amount attached to each noun. Amengual's definition of gender, however, is deficient: it is hard to understand, but above all it ignores features that new linguistic studies have considered of capital importance (p. 46): "Género viene del verbo latino *generare*, y asi el género significa substancia engendrada en una clase." ('Gender comes from the Latin verb *generare*, and thus means substance created in a class'). Today, the concept of gender is considered to be much more complex than that of number; twentieth-century linguistics has framed it in the more general notion of nominal classifiers. In nominal classifiers, gender is a residue which in English has already disappeared.

Although the only definition of gender found in the corpus is not well founded, all the grammars consulted offer two approaches to the issue: one based on sense, and the other based on the endings. Regarding the sense, they uphold the difference between men and male animals for the masculine gender vs. women and female animals for the feminine gender. Nobody refers to other differences for gender identified in later works referring to Catalan or to other languages with gender, such as the differences in the size of objects (such as *gaiato* / *gaiata* 'stick' / 'big stick' or *caragol* / *caragola* 'snail' / 'large snail') and in the pair 'tree-fruit'. This use already existed in Latin (*malus*/*mala*, *pomer* / *poma* 'apple tree' / 'apple') and survives in Castilian (*manzano* / *manzana* 'apple tree' / 'apple'), but not in Catalan.

As for the endings, grammarians develop very complex and very often tedious taxonomies. This facet cannot be analysed here; rather, we will focus on nouns with motion of gender, which can be either masculine or feminine depending on the sex of the being to which they refer.

Motion of gender is an idiosyncratic aspect of languages and dialects which allows them to be compared and contrasted. The grammars contain important information on this subject. The *Apuntes* by Nebot, a very perceptive scholar, is the most interesting. The information he provides is also found in Foulché-Delbosc and partly in Soler, but Nebot's text is the only one to include a particularly significant detail (p. 40): he notes that there are certain nouns with motion of gender, which, when used as adjectives (homonyms), abandon this feature and become invariable:

Table 5. Motion of gender in different grammars

Grammars	Nouns (masc.)	Nouns (fem.)	Adjectives (masc.)	Adjectives (fem.)
Soler			vigilant 'keeper'	vigilant
	cheneral 'general'	chenerala 'general woman'	cheneral 'general'	cheneral
Nebot	militar 'soldier'	militara	militar 'soldier'	militar
	chagant 'giant'	chaganta	chagant 'giant'	chagant
Foulché-Delbosc	general 'general'	generala	general 'general'	general

As regards the notion of number, the grammars include various aspects that merit attention. One especially interesting feature is the plural of nouns ending in an unstressed -*a*; this is a source of difference because in Eastern Catalan dialects -*as* and -*es* endings are neutralized in an unstressed schwa. The orthographic tradition represents this neutralization by the pair *dona*/*dones* 'woman/women', as does Nebot (because this was his pronunciation) along with Forteza, Foulché-Delbosc and Puiggarí. This example indicates that they were familiar with their orthographic tradition. In contrast, Amengual and Soler, both from the Balearic Islands, only show the result *dona*/*donas*, so they explain this case in terms of the general rule of plural formation. Both Forteza and Puiggarí include *donas*, but prioritize *dones*. However, Nebot distinguishes dones from words such as *tronada* 'thunderstorm', which is written *troná*: the general Valencian result of the loss of an intervocalic -*d*-, as in the ending -*ada*. He mentions the kind of word that forms the plural by adding -*es*: *tronaes*, *casaes* 'married women'.

The class of words represented by the pair *home* / *hòmens* 'man' / 'men' also receives very interesting explanations with regard to the expression of number. Neither Amengual, Soler nor Puiggarí take into account the plural *hòmens*, so they treat this case by applying the general rule for the plural: *home* / *homes*. Nebot is the only one to consider *home* / *hòmens*, since the Valencian variety has so far kept these plurals alive. Forteza mentions these two possibilities but notes that *homes* or *homos* are only found in Majorca. Therefore, he is familiar with the classical form, but does not mention the Valencian form. Foulché-Delbosc does not discuss *hòmens*, but, with reference to the diminutive, includes forms that belong to the *hòmens* group: *asenet* 'small donkey', *covenet* 'small basket', *Estevenet* 'young Steve', *homenet* 'small man', *jovenet* 'young man' and *orfenet* 'small orphan' (Cf. Nebot: *Hòme*, *orde* 'order', *òrgue* 'pipe organ', *òrfe*).

One final aspect of the *home* / *hòmens* group should also be mentioned. Grammars show a notable difference regarding the appreciative derivation (diminutive and augmentative, often with pejorative or derogatory values). All grammarians – except Soler, who does not mention the subject – consider that -*et*/-*eta* is a derivational suffix and -*às*/-*assa* is augmentative. In addition, -*ot*/-*ota* is considered as augmentative, with a derogatory or pejorative value, except by Puiggarí, which places -*ot*/-*ota* in the group of diminutives. The explanation for this difference is found on page 103 in Forteza's *Gramática*, where he explains that this Latin suffix has sometimes evolved with different values; he notes the example of -OTTUS/-OTTA, which in Catalan and Castilian has an augmentative-derogatory value but in Italian and French has acquired a diminutive sense. Influenced by French, Puiggarí defines *homenot* as a 'weak man'.

4. Conclusions

The study of the grammar corpus in this computerized form improves our understanding of linguistic variation in a period when dialectal studies were non-existent. In addition to particular linguistic features, these grammars highlight other areas of interest such as the importance of methodological aspects of the study of linguistics. This issue, and others, should be the subject of further analysis.

References

Alcover, A.M., & Moll, F. de B. (1930-1962). *Diccionari català-valencià-balear*. Palma de Mallorca: Moll [http://dcvb.iecat.net/].

Colón, G., & Soberanas, A.J. (1985). *Panorama de la lexicografia catalana*. Barcelona, Enciclopèdia Catalana.

Perea, M.P. (2003). Els dialectes catalans i el desenvolupament de la dialectologia durant el segle XIX. In *Actes del Dotzè Col·loqui Internacional de Llengua i Literatura Catalanes*. Universitat de París IV – Sorbonne 4-10 setembre 2000. Barcelona: AILCC, Publicacions de l'Abadia de Montserrat, 413-433.

Perea, M.P. (2008). Tomàs Forteza: capdavanter dels estudis lingüístics i gramaticals. In T. Forteza i Cortés. *Gramática de la lengua catalana*. vol. I, Barcelona: Publicacions de l'Abadia de Montserrat, 5-46

Perea, M.P. (2012a). Lexdialgram: un portal de diccionarios y gramáticas dialectales del catalán del siglo XIX. In *Actas del V Congreso Internacional de Lexicografía Hispánica*. Madrid: Universidad Carlos III (to be printed).

Perea, M.P. (2012b). Los dialectos catalanes en los diccionarios y gramáticas del siglo XIX. In *Actas del X Congreso Internacional de Lingüística General*, Universidad de Zaragoza, en prensa.

Pop, S. (1950). *La dialectologie: aperçu historique et méthodes d'enquêtes linguistiques*. Louvain: J. Duculot.